Ecological quality in freshwater streams is reflected across all three domains of life

Nadieh de Jonge[1], Franziska Kuntke[2], Martin Hesselsoe[3], and Jeppe Nielsen[2]

[1]Aalborg University
[2]Aalborg Universitet Teknisk-Naturvidenskabelige Fakultet
[3]NIRAS

April 28, 2020

Abstract

Regular assessment of surface waters to monitor ecological quality has become an increasing priority in recent years. Stream water quality assessment conventionally relies on taxonomic expertise, which can be costly and time-consuming. Next-generation sequencing approaches for high-throughput analyses of diverse ecosystems are increasingly used for environmental monitoring, and it has recently been shown that this methodology has a large potential for stream water quality assessments. In the present study we assess the potential of an amplicon sequencing based approach that simultaneously targets the 16S (prokaryotes) and 18S (eukaryotes) rRNA genes to capture a snapshot of the in the entire ecosystem composition. Results based on the analysis of 50 selected Danish streams showed that the domain-specific profiles can separate the samples into their respective ecological quality categories. Furthermore, it was possible to extract potential indicator organisms from all three domains associated to a single ecological status category, low and high quality ecosystems, as well as those with a high correlation to ecological status across all categories.

Introduction

The Water Framework Directive (WFD) outlines the requirements of the European Commission for regular assessment of European surface waters (Bengtsson et al., 2012), as a measure to monitor and ensure that water bodies of the member states have a high ecological quality. A large number of different protocols applicable to the assessment of water bodies by the member states are based on a diverse array of morphologically based descriptions of macroinvertebrates to plant life combined with chemical/physical parameters are currently implemented for assessment (Birk et al., 2012). This diversity in methodology has shown the need for cross comparison and standardisation to obey central directives. Furthermore, current assessment protocols rely heavily on invasive sampling and fauna identification by taxonomic experts, which is both time-consuming and costly (Leese et al., 2016). One example of stream ecosystem assessment protocols is the Danish Stream Fauna Index (DSFI) (Skriver, Friberg, & Kirkegaard, 2000). This method uses macroinvertebrate diversity as Biological Quality Elements (BQE) (Agouridis, Wesley, Sanderson, & Newton, 2015; Elbrecht & Leese, 2017). The water quality of the stream ecosystems are categorised into one of seven groups, ranging from poor (1), to very good (7) (Skriver et al., 2000). The DSFI categories translate to the ecological status classes as described by the WFD (Baattrup-Pedersen et al., 2004).

It has been proposed that bioassessment methods can be improved through the use of molecular techniques such as metabarcoding (Blackman et al., 2019; Pawlowski et al., 2018). Next generation sequencing (NGS) has become an increasingly mainstream and convenient approach to perform analysis of diverse ecosystems, The use of standardised DNA barcoding based on universal genetic markers allows for the identification of species through sequence data (Leese et al., 2016). Previous studies have yielded generalised protocols

1

for sampling and DNA extraction from marine sediments (Aylagas, Mendibil, Borja, & Rodríguez-Ezpeleta, 2016), as well as broad range primer sets targeting invertebrate biomarkers (Elbrecht & Leese, 2017), laying the foundation for high-throughput assay development. Recently, an amplicon sequencing approach based on targeting the cytochrome c oxidase I (*COI*) gene was developed and successfully applied as an alternative method of performing water quality assessments in streams (Kuntke, de Jonge, Hesselsøe, & Nielsen, 2020).

The drawback of molecular approaches for bioassessment is that design of barcoding targets can be problematic. Often not all of the desired diversity can be captured in a single barcode (i.e. one DNA amplification). Therefore, no consensus on genes and specific barcodes has been reached so far (Leese et al., 2016). Samples collected for fauna indexing according to water quality assessments contain DNA from all materials present within the ecosystem, including soil, water, plants and organisms inhabiting the sampling site. By choosing a universal approach that captures the majority of diversity present in a sample, it could potentially be possible to report on the state and dynamics of entire ecosystem, rather than focusing on a single source of DNA in the sample. Microbial communities of stream water ecosystems have previously been shown to correlate with land usage, and the status of the stream environment (Lear et al., 2013). Furthermore, quality assessment of ground and stream water via detection of selected microbial indicators such as *Escherichia coli* and *Clostridium perfringens* has also shown promising results (Francy, Helsel, & Nally, 2000). Moreover, the superior size of bacterial and archaeal diversities have previously been linked to changes in ecosystem quality after an oil spill (Urakawa, Garcia, Barreto, Molina, & Barreto, 2012). Collectively, these findings suggest that microbial community studies may present a promising approach for fast, accurate and cost effective bioassessment of freshwater ecosystems.

In the present study we characterise the complete biome profile of 53 bulk samples collected from freshwater streams across Denmark. The results are aligned with the pre-determined ecological quality status based on conventional water quality assessment using macroinvertebrate composition as BQE. A high-throughput amplicon sequencing approach using a universal primer set that targets all three domains (*Bacteria* ,*Archaea* and *Eukaryota*) was chosen for this purpose. The potential of whole biome analysis for ecosystem quality determination is assessed. Furthermore, taxa from all three domains with potential as indicators for ecological status were explored.

Material and methods

Traditional DSFI analysis and conversion to ecological status classes

Sampling of 53 Danish stream ecosystem sites using the kick-sampling method (Bradey & Ormerod, 2002) and subsequent DSFI analysis was performed by the Laboratory of Fish Ecology (Fiskeøkologisk Laboratorium, Helsingør, Denmark) in accordance with the Danish standardised DSFI protocol (Skriver, 1999). The samples used for the DSFI analysis was stored in 96 % ethanol at 4 °C until further analysis. After DSFI analysis, the scores were converted to ecological quality classes using a previously proposed distribution (Baattrup-Pedersen et al., 2004). An overview of all samples, their DSFI quality category and ecological quality class is shown in Table 1.

Sample homogenisation and DNA extraction

Prior to DNA extraction, samples were decanted using a mesh sieve (0.5 mm) and homogenised with a blender (JB 5160 BK, Braun GmbH, Germany) using 10 second cycles at speed setting 3. Total genomic DNA was extracted from 0.25 g of homogenised material using the QIAamp PowerFecal DNA kit (Qiagen, USA) according to manufacturer's recommendations. Extracted DNA was eluted to 50 μL elution buffer from the kit. Concentration of the extracted DNA was determined using Quant-IT Broad Range DNA assay kit (Thermo Fisher Scientific, USA) and an M200 Infinite PRO plate reader (TECAN, Switzerland).

Amplicon sequencing of SSU biomarkers

Molecular ecosystem profiles of the samples were captured using an amplicon sequencing approach targeting the genes coding for the small ribosomal subunits 16S (prokaryotes) and 18S (eukaryotes) simultaneously

using a previously described universal primer set: 926F (5'-AAACTYAAAKGAATTGACGG-3') and 1392R (5'-ACGGGCGGTGTGTRC-3') (Engelbrektson et al., 2010). Amplicon library PCR was performed on 10 ng of extracted DNA per 25 μL PCR reaction (1X Platinum High Fidelity buffer (Thermo Fisher Scientific, USA), 2 mU Platinum Taq DNA polymerase HF (Thermo Fisher Scientific, USA), 1.5 mM MgSO$_4$, 400 nM of each dNTP and 400 nM of each primer), using the following PCR programme: 2 minutes initial denaturation at 95 °C, 30 cycles of 20 seconds at 95 °C, 30 seconds at 52 °C, 60 seconds at 72 °C, final elongation for 5 minutes at 72 °C. PCR reactions were performed in duplicates, pooled afterwards and subsequently purified using AMPure XP bead protocol (Beckmann-Coulter, USA) using a sample:bead ratio of 5:4. Amplicon quality was assessed using TapeStation 2200 and D1000 ScreenTapes (Agilent, USA), and quantity was determined using Quant-IT High Sensitivity DNA assay kit (Thermo Fisher Scientific, USA). The amplicon libraries were subsequently barcoded in accordance with the Nextera XT barcoding protocol (Illumina, USA). Equimolar concentrations of the libraries were sequenced on a MiSeq platform using reagent kit v3 (2x300 PE) (Illumina, USA), with a final library pool concentration of 4 pM and 20 % PhiX spike-in.

Data processing and analysis

The obtained raw sequencing reads were processed using the AmpProc pipeline (v5.1) (https://github.com/eyashiro/AmpProc). Due to variable amplicon length between domains (the eukaryotic amplicons are too long to merge), the data was processed in single read mode without taxonomy assignment to generate a complete biome dataset of the reverse read (R1), and to extract the eukaryotic ZOTUs for individual domain analyses. A paired end treatment of the raw sequences was also performed to retrieve the highest possible taxonomic resolution for *Bacteria* and *Archaea* . Taxonomic classification was performed manually using SILVA S138 (Quast et al., 2013).

Data analysis was performed using R version 3.6.1 (R Development Core Team, 2020) wrapped by RStudio version 1.2.5001 (http://www.rstudio.com/). Visualisation of the obtained data and ordination analysis were performed using the packages ampvis2 (Andersen, Kirkegaard, Karst, & Albertsen, 2018) and ggplot2 (Wickham, 2016).

Results

Amplicon sequencing of 53 bulk samples from Danish freshwater streams yielded a total of 2,416,798 high quality sequence reads. The obtained sequencing depth was assessed using a rarefaction curve (Figure S1), and based on this, a minimum 10,000 sequences per sample was set as the criterion for inclusion in the analysis. Three samples were removed from the dataset as they did not meet this threshold, and 50 samples across all five ecological status categories entered the analysis with an average of 41,684 $\pm$ 7628 sequences per sample. A total of 13,446 ZOTUs were detected across all samples, with the majority representing the domain *Bacteria* (11,672 ZOTUs, 87.8 % of total reads, followed by *Eukaryota* (282 ZOTUs, 7.3 % of total reads) and *Archaea* (233 ZOTUs, 0.8 % of total reads). A total of 1259 ZOTUs (4.1 % of total reads) could not be classified and were removed from subsequent analyses.

Diversity and composition of molecular biome profiles in Danish freshwater streams

Diversity and distribution of the obtained biome profiles was assessed using the estimated richness as calculated per domain by alpha diversity index Chao1 (Figure 1). The lowest average number of bacterial ZOTUs was observed in the samples from good ecological status (4,243 $\pm$ 1,414 ZOTUs) (Figure 1a), while the highest was seen in the samples of poor ecological status (5,790 $\pm$ 1,777 ZOTUs). No immediate relationship between the bacterial diversity and ecological status of the sampled locations was apparent (p > 0.05), and a similar pattern of diversity was observed for the *Archaea* ZOTUs. A trend towards increased diversity with better ecological status was observed for the Eukaryotic ZOTUs (Figure 1b), where an average 54 $\pm$ 3 ZOTUs were observed in the samples representing bad ecological quality, opposite 93 $\pm$ 25 ZOTUs in samples of high ecological status. However, statistical testing revealed no significant differences (p > 0.05) in the estimated number of ZOTUs across all 5 ecological status categories in all three domains. *Bacterial* and *Eukaryotic* ZOTUs had a relatively low variation between samples of the same groups, compared to *Archaea* , where a larger variation was observed.

The composition of the molecular biome profiles across ecological status categories was further examined per domain (Figure S2-S4). Prior to analysis, all ZOTUs associated to Chloroplasts were filtered from the dataset. The majority of the abundant bacterial community members were observed across all ecological status groups (Figure S2), including multiple representatives of *Rhodoferax, Nitrospira, Comamonadaceae* and *Anaerolineaceae* . Representatives of the genus *Bacillus* , as well as one ZOTU representing *Nitrospira* were almost exclusively detected in samples representing good and high ecological quality. Representatives of the domain *Archaea* were detected sparsely in samples across all five ecological status categories (Figure S3). The majority of abundant archaeal ZOTUs were associated to the class *Bathyarchaeia* and the family *Nitrososphaeraceae* . For the *Eukaryota* , several abundant organisms were observed across all ecological status groups, including representatives of the insect order *Diptera* and the plant division *Magnoliophyta* (Figure S4). Other eukaryotic ZOTUs were more sparsely detected in the sampled locations, and included invertebrates, fungi and annelids. Representatives of the flatworm group *Continenticola* were primarily observed in samples of good and high ecological status.

Ecological status is reflected in combined and domain specific molecular biome profiles

The relationship between ecological status and the composition of the molecular biome profile was investigated through beta diversity analysis with ordination. Unconstrained ordination analysis using correspondence analysis (CA) showed little to no separation between ecological status groups (data not shown). However, the use of the ecological status of the samples as a constraint for canonical correspondence analysis (CCA) revealed strong clustering and separation of the sampled locations for the whole biome data, as well as the three individual domains (Figure 2). Near complete separation of the five ecological status groups was achieved for the whole biome data (Figure 2a), while the *Bacteria* (Figure 2b and *Eukaryota* (Figure 2d) also showed clustering by ecological status but with overlap between categories. The *Archaea* ZOTUs were not able to separate the sampled locations in a meaningful way (Figure 2c). The two lowest ecological statuses (Bad and Poor) were generally separated further away from the remaining samples, while overlap was seen primarily for locations of Good and High ecological status.

Potential quality indicators are present in all three domains of life

Ecological indices for quality assessment rely on indicator organisms associated to the criteria described in the individual methods. To explore potential indicator organisms across the three domains analysed in the present study, three different types of relationships were investigated. The whole biome and individual domain data were both subjected to same criteria to extract ZOTUs of interest as potential indicators. The results from the whole biome dataset were comparable to the combined results of the individual domains (data not shown). The individual domains were therefore selected for detailed analysis to provide the highest taxonomic resolution possible. A selection of the best potential indicators identified across the three domains of life is shown in Figure 3, the full result is displayed in Figure S5-S7.

The first type of indicators was assumed to be associated to a single ecological status category, and extracted by selecting all ZOTUs that were observed in one third of samples in a single category (with a minimum of n = 2). This yielded 54, 2, and 0 ZOTUs for *Bacteria* , *Archaea* and *Eukaryota* , respectively. ZOTUs representing *Rhodobacteraceae* , *Flavobacterium* and *Ferruginibacter* were associated to streams of bad and poor ecological status, while ZOTUs representing *Pirellulaceae* , *Comamonadaceae* and *Pseudonocardia* were associated to those of good and high ecological status (Figure 3a).

The second type of indicators was assumed to be associated to either low (bad-moderate, n = 12) or high (good-high, n = 38) quality ecosystems, under the definition of the WFD which states that all ecosystems must achieve an ecological status of good or higher. ZOTUs were extracted by selecting those that occurred in at least one third of samples in either group. This yielded 60, 0 and 3 ZOTUs for *Bacteria* , *Archaea* and *Eukaryota* , respectively. Representatives of *Ferribacterium* and *Rhodobacter* were found to be associated to streams of lower ecological status, while ZOTUs representing *Solibacillus* , *Sphingomonadaceae* , *Continenticola* and the insect order *Diptera* were associated with high quality ecosystems (Figure 3b).

The final type of indicator was assumed to have a strong correlation to ecological status across all categories.

ZOTUs that were observed in at least half of the samples in the dataset (n = 50) were extracted and a Spearman's correlation analysis was performed. A total of 546 ZOTUs met the criteria for *Bacteria* , and additional filtering for -0.5 < rho > 0.5 reduced this to 34 ZOTUs with strong correlation to ecological status. The *Archaea* and *Eukaryota* did not yield ZOTUs with a strong correlation, but yielded a total of 6 and 14 ZOTUs, respectively. Strong negative correlation between occurrence and ecological status was observed for ZOTUs representing OM60(NOR5) clade (r = -0.61), *Rhodobacteraceae* (r = -0.51) and *Caenogastropoda* (r = -0.47), while a strong positive correlatin was observed for ZOTUs representing *Calothrix* (r = 0.7), *Limnohabitans* (r = 0.78), *Nitrospira* (r = 0.71) and *Nitrosomonadaceae* (r = 0.65) (Figure 3c).

Discussion

The present study aimed to characterise the domain-specific diversity profiles of 53 Danish stream water sampling locations using an amplicon sequencing based method simultaneously targeting the bacterial, archaeal and eukaryotic composition. The purpose was to explore this method as a low cost and high throughput alternative to conventional stream-water quality assessments based on macroinvertebrates as BQE. The obtained domain-specific diversity profiles were able to separate the sampled streams based on their respective ecological status and revealed potential indicator organisms for further studies.

Biome analysis and composition of Danish freshwater streams

One aim of the present study was to assess the merit of using high-throughput technologies for stream water quality assessment to increase sensitivity and simultaneously reduce labour intensity and cost. One of the steps in conventional bioassessment protocols is pre-treatment and sorting of the collected samples. Previous studies have developed generalised protocols for sample collection and handling for this purpose (Aylagas et al., 2016). Recently obtained results from a study involving metabarcoding of stream water bodies showed that direct sample homogenisation without pre-treatment was equally effective for molecular analysis compared to samples that had been washed and sorted prior to DNA extraction (Kuntke et al., 2020). The obtained sequencing results in the present study supports the direct use of samples without pre-treatment, as it was possible to obtain high quality sequences to a sufficient depth for detailed biome analysis from all three domains in 50 out of 53 samples. The omission of pre-treatment steps from the methodology reduces the sample preparation time and the extent of the biases associated with this process.

The obtained sequencing data was of high quality and shown to be of sufficient sequencing depth for high resolution analysis of domain-specific community profiles (Figure S1), with the majority of the captured diversity originating from the bacterial communities present in the samples (Figure 1a). This was as expected, as sediment material makes up a large fraction of the sample, and the sediment microbiome is one of the most complex biomes that has been studied (Battin, Besemer, Bengtsson, Romani, & Packmann, 2016). The methodology of analysing multiple domains for ecosystem quality studies has previously been attempted in a study involving the use of multiple biomarkers (*16S rRNA* , *18S rRNA* and *COI* genes) and sequencing technologies aimed to develop a framework for metasystematic analysis of bulk samples containing arthropods (Gibson et al., 2014). The sequencing data obtained in the present study reveal a measured diversity for all domains that is multiple times higher than previously reported. Furthermore, the novel approach of using a single primer set to capture a comprehensive picture of the stream water biome simultaneously improves data handling and comparability, and highlights the quality and convenience of the chosen methodology.

A number of known sediment associated bacteria were detected abundantly across samples of all qualities (Figure S2-S4), including the genera *Rhodoferax* , *Nitrospira* and the family *Comamonadaceae* , as well as representatives of the archaeal family *Nitrosospaeraceae* (Battin et al., 2016). The composition of plant life in forest ecosystems has previously been used to predict soil microbiome profiles in grasslands, with moderate success (Leff et al., 2018), which highlights the relationship of the soil microbes with the flora and fauna of the ecosystem they are present in. Furthermore, microbial community monitoring has previously been applied in the tracking of bioremediation after oil spills in marine sediment environments (Acosta-González & Marqués, 2016; Urakawa et al., 2012). Moreover, the quality of the soil and its microbes has previously been shown to function as a predictor for ecosystem health based on land usage (Lear et al., 2013), and it can

thus be suggested that measuring the microbiome as well as the fauna and flora composition could improve assessment resolution for quality index studies in stream waters. Not surprisingly, a number of taxonomic groups containing organisms previously used as indicator species for bioassessment were also detected (Skriver et al., 2000), including representatives of *Diptera* , *Caenogastropoda* , *Mollusca* and *Coleoptera* .

Ecological status separates whole biome and individual domain compositions

It was not possible to separate the five ecological status groups based on the composition of the analysed streams on its own (data not shown). This was expected as the highly diverse and variable composition of the biome between samples obscures the relationship between ecological status and biological diversity. This observation is supported by previous studies exploring the potential of metabarcoding as an alternative for conventional bioassessments in freshwater streams (Kuntke et al., 2020). Another study exploring prediction of anthropogenic activity in rivers also found that the complete observed diversity could not explain ecosystem quality, and suggested the use of indicator organisms to specialise a potential model (Li et al., 2018).

To explore the relationship between ecological status and beta diversity further, a canonical correspondence model (CCA) was generated for the whole biome, as well as the individual domain data (Figure 2). Beta diversity analysis using CCA is a well-described and widely applied method in ecological studies (Braak & Verdonschot, 1995), which makes the chosen approach in the present study directly compatible with existing protocols for data analysis. The CCA model, constrained by ecological status, revealed that the whole biome data (Figure 2a) achieved the best separation, followed by the *Bacteria* (Figure 2b), where near complete separation of all five ecological status groups was achieved. Prokaryotic communities associated with sediments and surface waters have previously been shown to be sensitive to environmental changes, and have been suggested as a tool for biomonitoring of pollution (Li et al., 2018; Mlejnková & Sovová, 2010). The bacterial communities of freshwater streams may present a relatively unexplored approach with a high potential for the discovery of new indicators for bioassessment.

A gradient like overlap between streams of bad and poor, and moderate to high ecological status was observed for the eukaryotic data, which is in line with previous studies focusing on metabarcoding of invertebrates (Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017; Kuntke et al., 2020), as well as well-described ecological quality measurement protocols, which are based in the identification and abundance of chosen indicator species (Birk et al., 2012). The observed archaeal community was not able to separate the samples based on ecological quality in a meaningful way, however, this is likely related to the low presence and lack of differentiation to the surrounding environment and/or the coverage of the chosen primer set which might only capture a part of the archaeal taxa. It has previously been shown that archaeal communities in sediments are highly diverse, as well as sensitive to environmental change (Hoshino & Inagaki, 2019), and may be worth investigating in more details in relation to biomonitoring protocols of freshwater systems.

The domain-specific diversity analysis could potentially be extended with network analysis to reveal potential ecologically meaningful relationships within and across domains, which could strengthen the detection of indicator species and organisms associated to individual ecological status classes. A similar approach has previously been applied in paddy soils (Wang et al., 2017). Alternatively, indicator organisms could be extracted from the dataset to simplify the dimensionality of metabarcoding data and provide basis for a model describing the relationship between the biome and ecological status. This approach has previously been applied in rivers in China (Li et al., 2018). The data gathered in the present study is promising for this type of approach, however, due to the low number of samples with bad to moderate ecological status, the statistical strength of a predictive model generated based on the current dataset would be relatively low. Additional sampling to increase sampling size, especially in the lower quality ecosystems, would strengthen the data and enable the development of a predictive model based on biome composition data.

Potential quality indicators detected in all three domains of life

The potential of microbial community members as indicators of ecosystem quality has previously been shown in contaminated sediments and marine environments (Acosta-González & Marqués, 2016; Urakawa et al., 2012), as well as rivers (Li et al., 2018). The strongest correlation to ecological status among the individual

domains was observed in the ZOTUs from the bacterial community. Furthermore, correlations of interest from *Archaea* and *Eukaryotic* ZOTUs were also observed (Figure 3, S6 and S7)), but these were weaker compared to the *Bacteria* . The latter correlations may be affected by their relatively low abundance in the bulk samples, and may be more significant when implemented in a network based approach targeting cross-domain interactions as indicators of quality, due to the complex nature of stream ecosystems (Battin et al., 2016).

Bacterial communities have previously been shown to reflect environmental changes relating to land use (Lear et al., 2013), as well as (anthropogenic) pollution (Acosta-González & Marqués, 2016; Mlejnková & Sovová, 2010; Urakawa et al., 2012). Organisms previously reported in relation to ecosystem health were found among the potential indicator OTUs of the prokaryotes. The genus *Flavobacterium* was observed among the ZOTUs associated to lower quality streams, and has previously been linked to anthropogenic activity in surface waters (Acosta-González & Marqués, 2016; Mlejnková & Sovová, 2010). A positive correlation was observed between the ecological status and a representative of the family *Nitrosomonadaceae* and the genus*Nitrospira* . This correlation has previously been characterized in a marine ecosystem in which pollution was shown to induce significant changes in the nitrifying communities (Urakawa et al., 2012). Similar observations have described for increased abundances of the family*Rhodobacteraceae* exposed to oil pollution in sediments (Acosta-González & Marqués, 2016), however the indicator analysis in the present study yielded multiple ZOTUs representing this family, but with opposing tendencies. This highlights the importance of taxonomic resolution, as a family of bacteria can contain numerous of species with a wide range of functions and sensitivity to their environment.

Analysis of the microbial communities, and its vast diversity, present in stream bulk samples could provide an interesting supplement to existing metabarcoding approaches targeting biomarkers for eukaryotes such as the *COI* gene (Kuntke et al., 2020). Now, when applying a molecular approach for water quality assessment we are no longer limited to the composition of visible BQE´s, such as macroinvertebrates. The microbial community composition in the streams may often capture a better and more comprehensive and sensitive snapshot of the diversity in stream waters.

Complete biome capture as an alternative method for stream water quality studies

The metabarcoding approach applied in the present study is a potential cost effective and fast alternative to existing metagenomics-based analyses for biomonitoring as it targets specific genes selected to represent a given taxonomic affiliation, thereby capturing all three domains of life. The use of domain specific indicator genes facilitates detailed analysis covering the majority of ecosystem diversity, thereby exploiting the opportunity to capture previously undescribed taxonomic diversity. While the databases for ribosomal target gene sequences from*Bacteria* and *Archaea* are good, they remain relatively poor for eukaryotic organisms, especially within the invertebrates. Attempts to fill the gaps in reference databases are ongoing as part of national and international Barcode of Life projects (e.g. NorBOL, GBOL, and SwissBOL) (Jinbo, Kato, & Ito, 2011). However, biome profiling for fauna or flora indexing does not necessarily have to provide full taxonomic descriptions and affiliations. The results obtained in this study clearly show that members of all three domains reflect the ecological status of the sampled location. Furthermore, the results provide evidence for the presence of potential indicator organisms within all three domains of life, which may provide a potential for alternative water quality assessment.

In conclusion, simultaneous analysis of *Bacteria* , *Archaea*and *Eukaryota* using high-throughput amplicon sequencing was shown to be a promising method for convenient and fast assessment of stream water quality. Domain-specific biome profiles separated the samples from 50 water bodies with different geological and physiochemical environments into their respective pre-determined ecological status classes. The obtained data showed that it was possible to extract potential new indicator organisms from all three domains. The strongest correlation between ecological status and community composition was observed for the *Bacteria* .

Acknowledgements

7

References

Acosta-González, A., & Marqués, S. (2016). Bacterial diversity in oil-polluted marine coastal sediments. *Current Opinion in Biotechnology* , *38* , 24–32. doi: 10.1016/j.copbio.2015.12.010

Agouridis, C. T., Wesley, E. T., Sanderson, T. M., & Newton, B. L. (2015). *Aquatic Macroinvertebrates : Biological Indicators of Stream Health* . *5* , 1–5.

Andersen, K. S., Kirkegaard, R. H., Karst, S. M., & Albertsen, M. (2018). Ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *BioRxiv* , *299537* . doi: 10.1101/299537

Aylagas, E., Mendibil, I., Borja, A., & Rodriguez-Ezpeleta, N. (2016). Marine Sediment Sample Pre-processing for Macroinvertebrates Metabarcoding: Mechanical Enrichment and Homogenization. *Frontiers in Marine Science* , *3* (October), 1–8. doi: 10.3389/fmars.2016.00203

Baattrup-Pedersen, A., Friberg, N., Pedersen, M. L., Skriver, J., Kronvang, B., & Larsen, S. E. (2004). Anvendelse af Vandrammedirektivet i danske vandlob. Danmarks Miljoundersogelser. In *Faglig rapport fra DMU* . Retrieved from http://faglige-rapporter.dmu.dk

Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M., & Packmann, A. I. (2016). The ecology and biogeochemistry of stream biofilms. *Nature Reviews Microbiology* , *14* (4), 251–263. doi: 10.1038/nrmicro.2016.15

Bengtsson, L., Fortin, V., Gronewold, A. D., Tanino, Y., Surridge, B., Watson, N., . . . Herschy, R. W. (2012). Water Framework Directive. *Encyclopedia of Lakes and Reservoirs* , 872–876. doi: 10.1007/978-1-4020-4410-6_217

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., . . . Hering, D. (2012). Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators* , *18* , 31–41. doi: 10.1016/j.ecolind.2011.10.009

Blackman, R., Machler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., . . . Deiner, K. (2019). Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. *Metabarcoding and Metagenomics* , *3* , 49–57. doi: 10.3897/mbmg.3.34735

Braak, C. J. F. F., & Verdonschot, P. F. M. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences* , *57* (3), 255–289. doi: 10.1007/BF00877430

Bradey, D. C., & Ormerod, S. J. (2002). Evaluating the precision of kick-sampling in upland streams for assessments of long-term change: the effects of sampling effort, habitat and rarity. *Archiv Fur Hydrobiologie* , *155* (2), 199–221. doi: 10.1127/archiv-hydrobiol/155/2002/199

Elbrecht, V., & Leese, F. (2017). Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science* , *5* , 1–17. doi: 10.3389/fenvs.2017.00011

Elbrecht, V., Vamos, E. E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution* , *8* (10), 1265–1275. doi: 10.7287/PEERJ.PREPRINTS.2759V2

Engelbrektson, A., Kunin, V., Wrighton, K. C., Zvenigorodsky, N., Chen, F., Ochman, H., & Hugenholtz, P. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *The ISME Journal* , *4* (5), 642–647. doi: 10.1038/ismej.2009.153

Francy, D. S., Helsel, D. R., & Nally, R. A. (2000). Occurrence and distribution of microbiological indicators in groundwater and stream. *Water Environment Research* , *72* (2), 152–161. doi: 10.2307/25045352

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., . . . Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences of the United States of America* , *111* (22), 8007–8012. doi: 10.1073/pnas.1406468111

Hoshino, T., & Inagaki, F. (2019). Abundance and distribution of Archaea in the subseafloor sedimentary biosphere. *ISME Journal* ,*13* (1), 227–231. doi: 10.1038/s41396-018-0253-3

Jinbo, U., Kato, T., & Ito, M. (2011). Current progress in DNA barcoding and future implications for entomology. *Entomological Science* , *14* (2), 107–124. doi: 10.1111/j.1479-8298.2011.00449.x

Kuntke, F., de Jonge, N., Hesselsoe, M., & Nielsen, J. L. (2020). Stream water quality assessment by metabarcoding of invertebrates.*Ecological Indicators* , *111* , 105982. doi: https://doi.org/10.1016/j.ecolind.2019.105982

Lear, G., Washington, V., Neale, M., Case, B., Buckley, H., & Lewis, G. (2013). The biogeography of stream bacteria. *Global Ecology and Biogeography* , *22* (5), 544–554. doi: 10.1111/geb.12046

Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., . . . Zimmermann, J. (2016). DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe.*Research Ideas and Outcomes* , *2* , e11321. doi: 10.3897/rio.2.e11321

Leff, J. W., Bardgett, R. D., Wilkinson, A., Jackson, B. G., Pritchard, W. J., de Long, J. R., . . . Fierer, N. (2018). Predicting the structure of soil communities from plant community taxonomy, phylogeny, and traits. *ISME Journal* , 1–12. doi: 10.1038/s41396-018-0089-x

Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J., & Zhang, X. (2018). Application of Environmental DNA Metabarcoding for Predicting Anthropogenic Pollution in Rivers. *Environmental Science and Technology* , *52* (20), 11708–11719. doi: 10.1021/acs.est.8b03869

Mlejnkova, H., & Sovova, K. (2010). Impact of pollution and seasonal changes on microbial community structure in surface water. *Water Science and Technology* , *61* (11), 2787–2795. doi: 10.2166/wst.2010.080

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., . . . Kahlert, M. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of The Total Environment* , *637 –638* , 1295–1310. doi: 10.1016/j.scitotenv.2018.05.002

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* , *41* (D1), D590–D596. doi: 10.1093/nar/gks1219

R Development Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from https://www.r-project.org/

Skriver, J. (1999). Danish Stream Fauna Index (DSFI) as an indicator of rare and threatened benthic macroinvertebrates. In N. Friberg & J. D. Carl (Eds.), *Biodiversity in Benthic Ecology. Proceedings from Nordic Benthological Meeting in Silkeborg, Denmark, 13-14 November 1997* . National Environmental Research Institute, Denmark.

Skriver, J., Friberg, N., & Kirkegaard, J. (2000). Biological assessment of running waters in Denmark: introduction of the Danish Stream Fauna Index (DSFI). *Verh. Internat. Verein. Limnol.* ,*27* , 1822–1830.

Urakawa, H., Garcia, J. C., Barreto, P. D., Molina, G. A., & Barreto, J. C. (2012). A sensitive crude oil bioassay indicates that oil spills potentially induce a change of major nitrifying prokaryotes from the Archaea to the Bacteria. *Environmental Pollution* , *164* , 42–45. doi: 10.1016/j.envpol.2012.01.009

Wang, H., Wei, Z., Mei, L., Gu, J., Yin, S., Faust, K., . . . Yin, S. (2017). Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. *Soil Biology and Biochemistry* , *105* , 227–235. doi: 10.1016/j.soilbio.2016.11.029

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* . Springer-Verlag New York.

Data accessibility

All raw sequencing data has been made available at the European Nucleotide Archive (ENA) under the project accession number PRJEB37542.

Author contributions

MH and JLN conceived the ideas, designed the work and achieved the funding, FK did the sampling and laboratory work, NdJ analysed the data with guidance from JLN. NdJ and JLN led the writing of the manuscript with contributions from MH and FK.

Competing interests

The authors declare that they have no competing interests.

Tables

Table 1: Overview of samples, DSFI quality and ecological status.

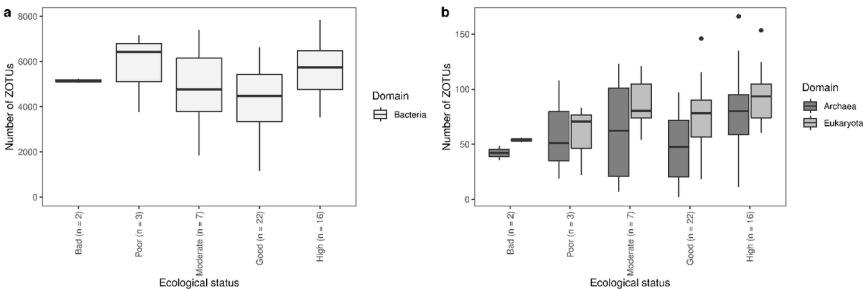| DSFI category	Ecological status	n
1-2	Bad	2
3	Poor	3
4	Moderate	7
5-6	Good	22
7	High	16

Figures

Figure 1: Estimated richness (ChaoI) of the molecular biome profiles from 50 samples is shown as boxplots for *Bacteria* (a) and *Archaea* and *Eukaryota* (b), sorted by ecological status and divided into the three detected domains.
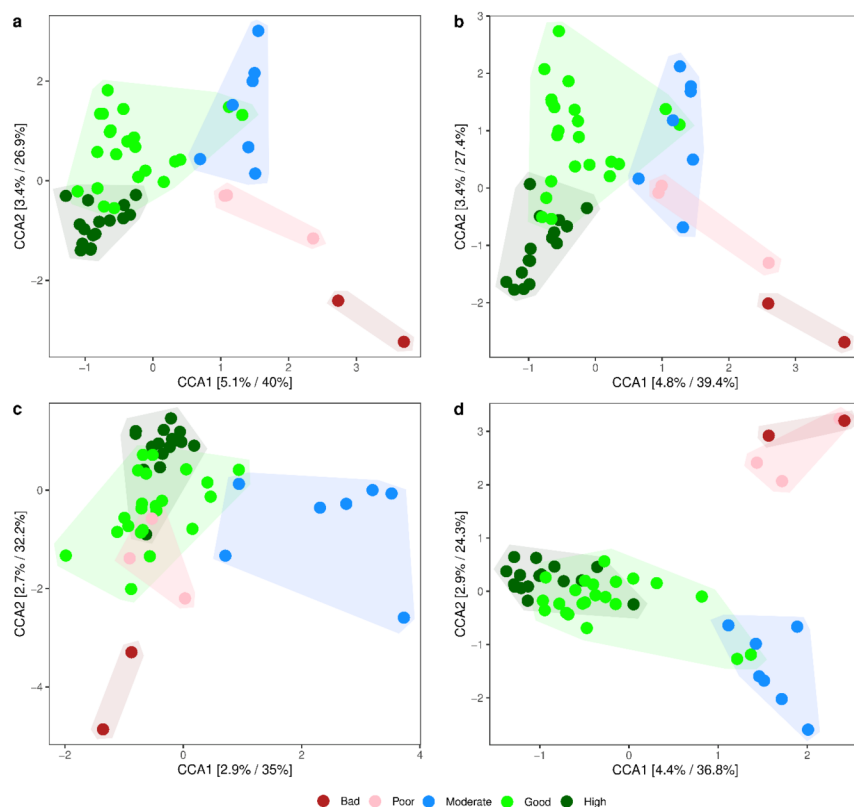
Figure 2: Canonical correspondence analysis of freshwater stream biomes, constrained by ecological status, for bacteria (a), archaea (b), and eukaryotes (c). Samples are coloured by ecological status, and a polygon is drawn around locations with the same ecological status.
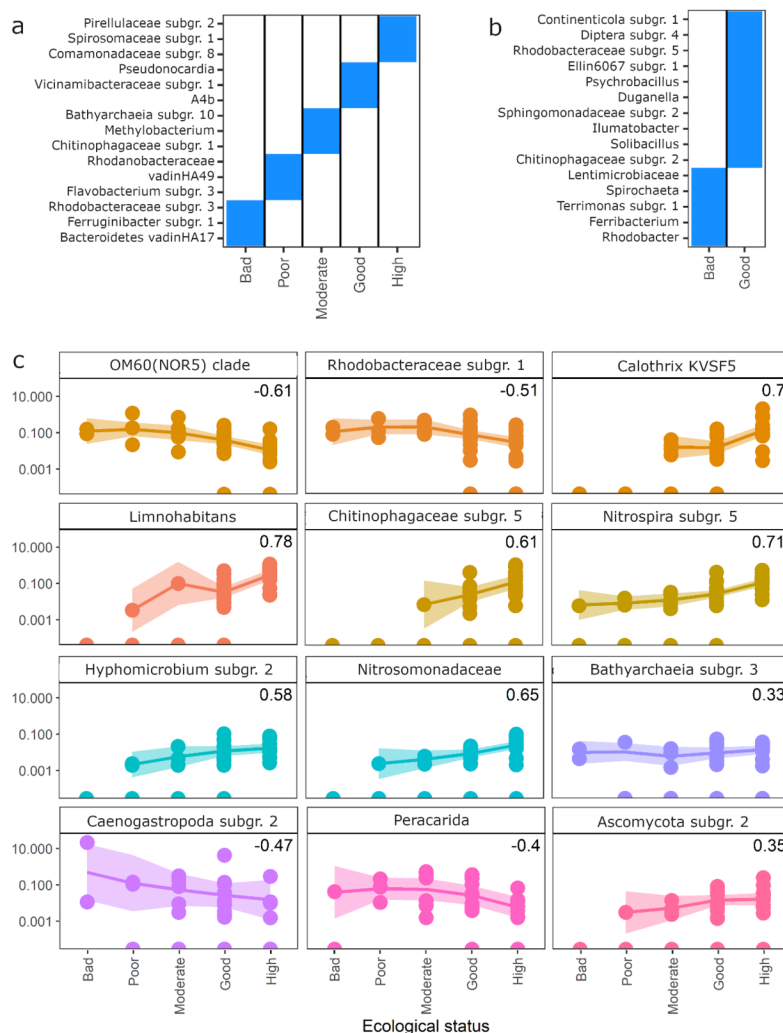
Figure 3: Relationship between ecological status and occurrence of organisms from all three domains of life. ZOTUs from Bacteria, Archaea and Eukaryota with association to a single category (a), low or high quality ecosystems (b) or strong correlation to ecological status (c) were extracted from the dataset. A selection of the most influential ZOTUs for all three domains is displayed as heatmaps showing presence/absence and scatterplots with loess lines and 95 % confidence intervals with Spearman's rho displayed in each panel. The complete result of the analysis is shown in Figures S5-S7.